27ª Semana Olímpica

Bento Gonçalves - RS - Janeiro de 2023

Amigo Secreto, Escolha Ótima, Agulha de Buffon e Polinômios Aleatórios

Ricardo Misturini
ricardo.misturini@ufrgs.br

Nesta aula discutiremos quatro problemas de probabilidade:

Problema 1. (Amigo Secreto) Na brincadeira do amigo secreto com N pessoas, cada uma escreve seu nome em um pedaço de papel e o deposita em um recipiente. Em seguida, cada pessoa, ao acaso, pega um dos pedaços de papel. Qual é a probabilidade de ninguém pegar seu próprio nome?

Problema 2. (Escolha Ótima) Você tem a tarefa de contratar uma pessoa para uma vaga de emprego, para a qual há N candidatos. Você não conhece os candidatos, mas sabe que seria possível ranqueá-los do melhor para o pior se todos fossem entrevistados. Os candidatos aparecem para serem entrevistados sequencialmente, em uma ordem aleatória. Depois de cada entrevista, você deve decidir entre as opões (A) ou (B):

(A) Rejeitar o candidato definitivamente, e entrevistar o próximo.

(B) Contratar o candidato, sem entrevistar mais ninguém.

O problema consiste em obter uma estratégia que maximize a probabilidade de que o melhor candidato seja contratado.

Problema 3. (Agulha de Buffon) Suponha que temos um piso formado por tábuas paralelas, todas de mesma largura e que jogamos uma agulha sobre o piso. Qual é a probabilidade de que essa agulha cruze uma das linhas entre as duas tábuas?

Problema 4. (Polinômio Aleatório) Seja $p(x) = X_0 + X_1x + X_2x^2 + \cdots + X_nx^n$ um polinômio aleatório de grau n, onde $X_0, X_1, \ldots X_n$ são variáveis aleatórias independentes com distribuição normal padrão. Seja N_n o número de raízes reais desse polinômio, e seja $\mathbb{E}(N_n)$ o valor esperado de N_n. O problema consiste em obter uma expressão para $\mathbb{E}(N_n)$.

Surpreendentemente, os Problemas 1 e 2 se relacionam com número de Euler, $e \approx 2{,}718$ (mais especificamente, com o número $1/e \approx 36{,}79\%$). Além disso os Problemas 3 e 4 se relacionam entre si e com o número π.

O número e

O número e é uma constante matemática aproximadamente igual a 2,71828 que pode ser caracterizada de várias maneiras. A expressão $\left(1+\frac{1}{n}\right)^n$ surge naturalmente no cálculo de alguns juros compostos. Por exemplo, um valor que sofre dois aumentos consecutivos de 50% é multiplicado por $(1+\frac{1}{2})^2 = 2{,}25$, de modo que o aumento total é de 125%. Um valor que sofre 10 aumentos consecutivos de 10% é multiplicado por $(1+\frac{1}{10})^{10} \approx 2,59$.

O número e é definido como o limite

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \approx 2,71828$$

É possível mostrar que e também é o valor da soma infinita $e = \sum_{k=0}^{\infty}\frac{1}{k!}$ o que é uma consequência da expansão da função exponencial $f(x) = e^x$ em série de potências $e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}$. E, fazendo $x = -1$, obtemos também que

$$\frac{1}{e} = \sum_{k=0}^{\infty}\frac{(-1)^k}{k!} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \cdots \approx 0,367879.$$

O problema do amigo secreto

Voltemos ao Problema 1. Na brincadeira do amigo secreto, quando alguém sorteia o próprio nome, dizemos que o sorteio não deu certo e é preciso refazer o sorteio. Queremos calcular a probabilidade de que ninguém sorteie o próprio nome, isto é, a probabilidade de que o sorteio funcione. A resposta exata depende de N, o número de participantes.

Vamos pensar (sem fazer contas) em uma pergunta relacionada para testar a nossa intuição. Como função de N, o que será que acontece com essa probabilidade quando N cresce? A probabilidade aumenta? Diminui? Ou tem um comportamento não monótono? Por exemplo, será mais fácil ninguém pegar o próprio nome quando a brincadeira é feita com um grupo de 20 amigos ou quando a fazemos com todos toda a população mundial (cerca de 8 bilhões)? A resposta não é óbvia. Na segunda situação, se olharmos para uma pessoa específica, a probabilidade de que ela sorteie o próprio nome é extremamente baixa. Por outro lado, como há muitas pessoas na brincadeira e cada uma tem um pouco de chance de sortear o próprio nome, então pode ser que de alguma maneira essas pequenas probabilidades se unam de modo que a chance de que pelo menos uma sorteie o próprio nome não seja tão baixa. Mostraremos na aula que nas duas situações ($N = 20$ ou $N = 8\cdot 10^9$) a probabilidade do sorteio funcionar é praticamente a mesma.

Mais precisamente, mostraremos que se a brincadeira é feita com N pessoas, a probabilidade de que ninguém sorteie o próprio nome é dada por

$$\sum_{k=0}^{N}\frac{(-1)^k}{k!} = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \frac{1}{5!} + \cdots + \frac{(-1)^N}{N!}.$$

Essa é a série de $1/e$ truncada. Quando N cresce esse número converge para $1/e \approx 36,79\%$. Essa convergência é muito rápida, para $N = 5$ a probabilidade já está na casa dos 36%.

O problema da escolha ótima

No Problema 2 assume-se que todas as permutações possíveis dos N candidatos são equiprováveis e portanto o melhor candidato pode estar em qualquer posição com igual probabilidade. Vamos dizer que obtemos sucesso se selecionarmos o melhor candidato.

Um exemplo de estratégia (que não é boa) seria simplesmente selecionar o primeiro candidato entrevistado. Nesse caso a probabilidade de sucesso é $1/N$, que é muito pequena se N é grande. Será que existe alguma estratégia tal que a probabilidade de sucesso não tende a zero quando N cresce?

Uma ideia um pouco melhor que a anterior seria rejeitar, digamos, a primeira metade dos candidatos e a partir daí selecionar o primeiro candidato que for melhor que todos os anteriores, caso esse candidato exista (caso contrário, fracassamos). Uma condição suficiente para o sucesso seguindo essa estratégia é que o melhor candidato esteja na segunda metade e o segundo melhor esteja na primeira metade da fila dos candidatos. Assim vemos que a probabilidade de sucesso seguindo essa estratégia é maior que 25%.

A estratégia acima parece boa, mas rejeitar os primeiros $N/2$ candidatos foi uma escolha arbitrária. É possível mostrar que a estratégia ótima é da seguinte forma: rejeitar os primeiros n candidatos e, a partir daí selecionar o primeiro candidato que for melhor que todos os anteriores. Na aula vamos obter uma expressão explícita em função de n e N para a probabilidade de sucesso seguindo essa estratégia e mostraremos que essa probabilidade é a maior possível quando $n \approx N/e$, isto é, aproximadamente 36,78% dos candidatos. Mais precisamente, mostraremos que o n ótimo é tal que

$$\frac{N}{e} - 1 < n < \frac{N}{e} + 1.$$

Além disso, seguindo essa estratégia a probabilidade de sucesso é superior a $1/e \approx 36{,}78\%$ e converge para $1/e$ quando N cresce.

A agulha de Buffon

O Problema 3 foi proposto e resolvido no século XVIII pelo Francês Georges-Louis Leclerc, conhecido como conde de Buffon. Conforme veremos na aula, sendo d o espaçamento entre as tábuas e ℓ o comprimento da agulha, $\ell < d$, a probabilidade de que ocorra um cruzamento é dada por $p = \frac{2}{\pi}\frac{\ell}{d}$. Dessa forma, em princípio, poderíamos obter uma aproximação para π repetindo esse experimento muitas vezes e observando a proporção dos casos em que ocorre cruzamento (isso é a essência do moderno Método de Monte Carlo). Discutiremos um pouco da história desse "método" para calcular π. História essa que elenca aproximações supostamente obtidas com esse experimento físico, que são suspeitas por serem "boas demais".

Na aula apresentaremos duas demonstrações do resultado de Buffon: 1) a prova clássica, que envolve um pouco de trigonometria básica e o cálculo de uma integral. 2) a prova elegante, de Barbier (1860), a qual consegue simplificar o problema jogando "agulhas tortas".

Surpreendentemente a ideia da prova de Barbier pode ser usada no Problema 4, para obter uma representação simples para $\mathbb{E}(N_n)$ em termos de uma integral.

Raízes reais de polinômios aleatórios

Existe uma vasta literatura relacionada ao Problema 4. Um dos primeiros resultados neste contexto foi obtido por Bloch & Pólya (1936), que estudaram o caso em que os coeficientes X_k são independentes e uniformemente distribuídos em $\{-1, 0, 1\}$ e obtiveram que $\mathbb{E}(N_n) \ll \sqrt{n}$, onde a notação $f(n) \ll g(n)$ significa que existe uma constante C tal que $f(n) \leq Cg(n)$ para todo n. Essa cota não é ótima. Em uma série de artigos iniciada em 1938, Litlewood e Offord provaram para várias distribuições (tais como, uniforme em $\{-1, 1\}$, Gaussiana centrada e uniforme em $[-1, 1]$) que $N_n \ll \log^2(n)$, com alta probabilidade, isto é, com probabilidade tendendo a 1 quando $n \to \infty$. Em 1943, Kac obteve uma fórmula exata para $\mathbb{E}(N_n)$, no caso em que os coeficientes X_k são Gaussianas, e com isso obteve a estimativa $\mathbb{E}(N_n) \sim \frac{2}{\pi}\log(n)$, onde a notação $f(n) \sim g(n)$ significa que $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

Resultados análogos para outras distribuições se tornaram um grande desafio. Por exemplo, em 1956, Erdös e Offord estenderam o resultado de Kac para o caso em que os coeficientes X_k são uniformes em $\{-1, 1\}$.

Em 1995, Edelman & Kostlan abordaram o problema de um ponto de vista geométrico. Eles observaram, no caso Gaussiano, que o número de raízes reais do polinômio $p(t)$ pode ser interpretado como o número de interseções de uma certa curva $\{\gamma(t)\}_{-\infty<t<\infty}$ contida na esfera unitária $S^n \subseteq \mathbb{R}^{n+1}$ com um equador aleatório dessa esfera. Dessa forma, o problema passa a ser o cálculo do valor esperado do número de interseções. Isso é uma versão do problema de Buffon na esfera de dimensão n. Podemos usar a mesma estratégia de Barbier para concluir que esse valor esperado é proporcional ao comprimento da curva γ e também obter a constante de proporcionalidade, que é $1/\pi$. Assim é possível obter uma fórmula exata para o número esperado de raízes reais:

$$\mathbb{E}(N_n) = \frac{1}{\pi}\int_{-\infty}^{\infty} \|\gamma'(t)\| dt.$$

Em nossa aula não estaremos interessados nos resultados assintóticos decorrentes dessa representação, mas mostraremos os detalhes de como ela pode ser obtida, e como o mesmo argumento pode ser utilizado em outras classes de funções aleatórias, como é o caso da série de Dirichlet aleatória.

Referências

[1] M. Aigner, G. M. Ziegler. *Proofs from the book*. 6th Edition. Springer (2018).

[2] M. Aymone, S. Frometa, R. Misturini. *How many real zeros does a random Dirichlet series have?* Eletronic Journal of Probabiliy, 29 (2024).

[3] M. Babović, V.Babović. *How needless are Buffon's needles?* Eur. J. Phys. 34 (2013)

[4] L. Badger. *Lazzarini's lucky approximation of* π. Mathematics Magazine. Vol. 67, No. 2 (1994).

[5] A. Bloch, G. Pólya, *On the roots of certain algebraic equations*, Proc. London Math. Soc. 33(1932), 102–114.

[6] A. Edelman and E. Kostlan, *How many zeros of a random polynomial are real?*, Bull. Amer. Math. Soc. (N.S.), 32 (1995), pp. 1–37.

[7] P. Erdös and A. C. Offord, *On the number of real roots of a random algebraic equation*, Proc. London Math. Soc. (3), 6 (1956), pp. 139–160.

[8] T. S. Ferguson. *Who solved the secretary problem?* Statistical Science, Vol. 4(3) (1989)

[9] J. Gilbert and F. Mosteller. *Recognizing the Maximum of a Sequence*. Journal of the American Statistical Association. Vol. 61(313) (1966)

[10] M. Kac, *On the average number of real roots of a random algebraic equation*, Bull. Amer. Math. Soc., 49 (1943), pp. 314–320.

[11] C. Landim. *Otimização Estocástica*. 18º Colóquio Brasileiro de Matemática, IMPA (1991). https://impa.br/wp-content/uploads/2017/04/18_CBM_91_04.pdf

[12] M. Lazzarini *Un'applicazione del calcolo della probabilità alla ricerca sperimentale di un valore approssimato di* π. Periodico di Matematica per l' insegnamento secundario 4 140–143. (1901)

[13] J. E. Littlewood and A. C. Offord, *On the Number of Real Roots of a Random Algebraic Equation*, J. London Math. Soc., 13 (1938), pp. 288–295.

[14] ——, *On the number of real roots of a random algebraic equation. III*, Rec. Math. [Mat. Sbornik] N.S., 12(54) (1943), pp. 277–286.

[15] R. Misturini. *Probabilidade: Primeiras noções e alguns problemas - Aula 03*. Prolímpico - 5ª Edição. IMPA (2022). https://www.youtube.com/watch?v=Of63aMN_hc4

[16] C. G. Moreira. *Amigo oculto*. RPM 15 (1989). https://w3.impa.br/~gugu/amigooculto.pdf

[17] T. Tao, V. Vu. *Local Universality of Zeroes of Random Polynomials*. International Mathematics Research Notices (2014).